DOCUMENT RESUME

ED 453 263 TM 032 800

AUTHOR McMillan, James H.; Workman, Daryl J.
TITLE Classroom Assessment and Grading Practices: A Review of the Literature.
INSTITUTION Metropolitan Educational Research Consortium, Richmond, VA.
PUB DATE 1998-05-00
NOTE 44p.
PUB TYPE Information Analyses (070)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Classroom Techniques; Educational Research; Elementary Secondary Education; *Evaluation Methods; Grades (Scholastic); *Grading; *Instructional Effectiveness; Literature Reviews; *Student Evaluation

ABSTRACT

This review of literature is an analysis of completed research on the nature and effect of classroom assessment practices and grading. An examination of the literature in this field suggests that it is divided into these categories: (1) definitions of classroom assessment; (2) classroom assessment practices; (3) grading practices; and (4) the effect of classroom assessment and grading practices on student learning and motivation. The review demonstrates that there is little empirical evidence of the specific effects of using particular assessments and grading practices. This results in part from the complex nature of teaching. Several newer forms of assessment, such as performance-based assessment and portfolio assessments, are based on recent research on cognitive learning, but they are based on theory and not empirical evidence. Some suggestions are made for combining different classroom assessment components into a framework that makes sense for explaining how classroom assessment contributes to student motivation and achievement. An annotated bibliography of 20 sources is included. (Contains 30 references.) (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 453 263

CLASSROOM ASSESSMENT AND GRADING PRACTICES:

A REVIEW OF THE LITERATURE

METROPOLITAN EDUCATIONAL RESEARCH CONSORTIUM

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

G. D. Hipp

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

CHESTERFIELD COUNTY PUBLIC SCHOOLS • COLONIAL HEIGHTS CITY SCHOOLS • HANOVER COUNTY PUBLIC SCHOOLS • HENRICO COUNTY PUBLIC SCHOOLS • HOPEWELL CITY PUBLIC SCHOOLS • POWHATAN COUNTY PUBLIC SCHOOLS • RICHMOND CITY PUBLIC SCHOOLS • VIRGINIA COMMONWEALTH UNIVERSITY

TM032800

BEST COPY AVAILABLE

METROPOLITAN EDUCATIONAL RESEARCH CONSORTIUM

PO BOX 842020 • RICHMOND VA 23284-2020 • Phone: (804) 828-0478
FAX: (804) 828-0479 • E-MAIL: jmcmilla@saturn.vcu.edu

MERC MEMBERSHIP
James H. McMillan, Director

CHESTERFIELD COUNTY PUBLIC SCHOOLS
William C. Bosher, Jr., Superintendent

COLONIAL HEIGHTS CITY SCHOOLS
James L. Ruffa, Superintendent

HANOVER COUNTY PUBLIC SCHOOLS
Steward D. Roberson, Superintendent
Chairman, MERC Policy & Planning Council

HENRICO COUNTY PUBLIC SCHOOLS
Mark A. Edwards, Superintendent

HOPEWELL CITY PUBLIC SCHOOLS
David C. Stuckwisch, Superintendent

POWHATAN COUNTY PUBLIC SCHOOLS
Margaret S. Meara, Superintendent

RICHMOND CITY PUBLIC SCHOOLS
Albert J. Williams, Superintendent

VIRGINIA COMMONWEALTH UNIVERSITY
John S. Oehler, Dean
School of Education

Virginia Commonwealth University and the school divisions of Chesterfield, Colonial Heights, Hanover, Henrico, Hopewell and Richmond established the **Metropolitan Educational Research Consortium** (MERC) on August 29, 1991. The founding members created **MERC** to provide timely information to help resolve educational problems identified by practicing professional educators. **MERC** membership is open to all metropolitan-type school divisions. It currently provides services to 9,000 teachers and 138,000 students. **MERC** has base funding from its membership. Its study teams are composed of University investigators and practitioners from the membership.

MERC is organized to serve the interests of its members by providing tangible material support to enhance the practice of educational leadership and the improvement of teaching and learning in metropolitan educational settings. **MERC's** research and development agenda is built around four goals:

- To improve educational decision-making through joint development of practice-driven research questions, design and dissemination,
- To anticipate important educational issues and provide leadership in school improvement,
- To identify proven strategies for resolving instruction, management, policy and planning issues facing public education, and
- To enhance the dissemination of effective school practices.

In addition to conducting research as described above, **MERC** will conduct technical and issue seminars and publish reports and briefs on a variety of educational issues.

CLASSROOM ASSESSMENT AND GRADING PRACTICES:

A REVIEW OF THE LITERATURE

Prepared by:
James H. McMillan
Professor
Educational Studies

Daryl J. Workman
Research Fellow

Virginia Commonwealth University
May, 1998

Copyright© 1998. Metropolitan Educational Research Consortium (MERC), Virginia Commonwealth University.

* The views expressed in MERC publications are those of individual authors and not necessarily those of the Consortium or its members.

Table of Contents

Review of Literature

Introduction 1

I. What is Classroom Assessment? 3

II. Classroom Assessment Practices 5

III. Grading Practices 10

IV. Effect of Classroom Assessment and Grading Practices on Student Learning and Motivation 16

Summary and Implications 23

References 28

Annotated Bibliography 31

CLASSROOM ASSESSMENT AND GRADING PRACTICES:

A REVIEW OF THE LITERATURE

INTRODUCTION

This review of literature is an analysis of completed research on the nature and effect of classroom assessment practices and grading. In recent years the assessment of student performance has become a central focus of efforts to reform education (Cizek, 1997). Policy-makers have increasingly seen assessment as a measure of student and school accountability, influencing curriculum and teaching. At the center of this movement is the classroom teacher. It is the teacher who communicates standards and expectations through the assessments students experience, and it is the teacher who makes decisions daily about what students learn. Classroom assessments, because students experience them continuously, are what have meaning to students concerning their abilities and achievement. Competent teachers use assessment to inform their instruction and determine student strengths and weaknesses.

The revived interest in assessment has resulted in part by advances in cognitive learning theory, motivation, and constructivist learning. These fields have shown that effective instruction does much more than simply present information to students. Rather, good instruction provides an environment that engages students in active learning that connects new information with existing information. Learning is an ongoing, self-regulated process in which students actively receive, interpret, and relate information in a meaningful way to what they already know and understand. Recent motivational research has suggested that specific and meaningful feedback to students help determine student self-efficacy and self-confidence (Brookhart, 1997).

Effective assessment is consistent with these new findings concerning student learning and motivation. In the past decade some clear trends have emerged in classroom assessment. Established practice of using objective assessments at the end of instruction are being supplemented with what are called "alternative" assessments, given during as well as at the end of instruction. Alternative assessments include authentic assessments, performance-based assessments, portfolios, exhibitions, journals, reflections, demonstrations, and other forms of assessment that require the active construction of meaning rather than passive regurgitation of isolated facts. The "new" assessments require students to be engaged in thinking skills and problem solving. These and other recent trends in classroom assessments are summarized in Table 1.

Table 1

Recent Trends in Classroom Assessment[1]

FROM	TO
Sole emphasis on outcomes	Assessing process
Isolated skills	Integrated skills
Isolated facts	Application of knowledge
Paper-and-pencil tasks	Authentic tasks
Decontextualized tasks	Contextualized tasks
A single correct answer	Many correct answers
Secret standards	Public standards
Secret criteria	Public criteria
Individuals	Groups
After instruction	During instruction
Little feedback	Considerable feedback
"Objective" tests	Performance-based tests
Standardized tests	Informal tests
External evaluation	Student self-evaluation
Single assessments	Multiple assessments
Sporadic	Continual
Conclusive	Recursive

[1] McMillan, 1997, p. 15

Despite the growing importance of classroom assessment and the introduction of new methods of assessment, there is relatively little research on the nature and effects of classroom assessments on student learning and motivation (Stiggins, 1997). Most assessment research has focused on standardized testing, despite evidence that teachers spend considerable time assessing students, and that student well-being is influenced by the quality of assessments given by the teacher (Stiggins and Conklin, 1992). Also, there is little empirical research on classroom assessments, with measurement experts tending instead to pay much more attention to large scale testing than classroom assessment. It is also evident that many teachers lack assessment competency (Plake and Impara, 1997). This isn't too surprising, however, since less than 50% of the teacher certification programs in the United States require no measurement course (Schafer, 1993). This remains the case, despite the fact that teacher standards for assessment competency were identified in 1990 (AFT, NCME, NEA, 1990).

In examining the classroom assessment and grading literature the research seems to be divided into four categories: I. definitions of classroom assessment, II. classroom assessment practices, III. grading practices, and IV. the effect of classroom assessment and grading practices on student learning and motivation. The following review of literature is organized according to these four categories.

I. What is Classroom Assessment?

Given the recent use of the general term "assessment", it is important to clarify what is meant by "classroom assessment." According to Cizek (1997), there are four definitions of assessment.

The term can refer to formats for gathering information, such as using a portfolio or performance assessment. Some see assessment as referring to a new attitude toward the way students are tested – away from standardized multiple choice. The term has come to represent a new ethos of empowerment to hold students and schools accountable. Finally, assessment can refer to a new process of gathering, synthesizing, and using information, one that is similar to what doctors and psychologists use when diagnosing and treating patients. These connotations suggest a much broader definition than what is typically conveyed when using the term "test."

In the context of teaching, this more general notion is represented by contemporary definitions of classroom assessments:

> [Classroom assessment is] the planned process of gathering and synthesizing information relevant to the purposes of (a) discovering and documenting students' strengths and weaknesses, (b) planning and enhancing instruction, or (c) evaluating progress and making decisions about students. (Cizek, 1997, p. 10)
>
> [Classroom assessment is] the collection, synthesis, and interpretation of information to aid the teacher in decision making. (Airasian, 1997, p. 4)
>
> [Classroom] assessment is a formal attempt to determine students' status with respect to diagnosing students' strengths and weaknesses, monitoring students' progress, assigning grades to students, and determining instructional effectiveness. (Popham, 1995, p. 3, 7)
>
> Classroom assessment can be defined as the collection, interpretation, and use of information to help teachers make better decisions. (McMillan, 1997, p. 8)

It is evident that these definitions provide a broad descriptor for what teachers must do. The term is clearly not the same as "test," "measurement," or "evaluation." A test is a single type of assessment in which students answer questions in a paper-and-pencil format, such as a multiple choice matching, or short answer test. End of unit, final exams, and pop quizzes are familiar types of tests. Measurement has traditionally been defined as a systematic process of assigning

numbers to performance. These numbers are used to differentiate degrees of a trait, characteristic, or behavior. This process can be quantitative or qualitative. Evaluation is making judgments about the quality of something, an interpretation of the results obtained from a test or some type of measurement to know what the results mean. For example, a score of 80% correct may be interpreted to mean that a student has mastered a skill. These evaluations are the decision making aspect of classroom assessments. Such decisions range from giving grades to knowing the focus of subsequent instruction.

For this review, then, classroom assessment is defined as the collection, synthesis, interpretation, and use of information to aid teacher decision making. Classroom assessment begins with the identification of a purpose for gathering the information, proceeds to selection of an appropriate way to gather information, and concludes with use of the results to enhance the quality of teachers' decisions.

II. Classroom Assessment Practices

Prior to the mid 1980s the literature on educational assessment focused almost exclusively on large-scale standardized testing. According to Stiggins and Conklin (1992), most inquiry on classroom assessment was based on a conceptualization similar to what had been developed for standardized testing, emphasizing paper and pencil, multiple choice testing. Furthermore, the only written standards for assessment, *Standards for Educational and Psychological Testing*, dealt primarily with standardized tests. Finally, during the 1980s the emerging literature about teacher decision-making, teacher behavior, and student achievement found little on how classroom assessments relate to teaching or learning. Shulman (1980) concluded that most of the

paper and pencil tests used for assessment were inconsistent with, and often irrelevant to, the realities of teaching. Haertel, et al. (1984), in a review of research on high school testing, concluded that little is known about teachers' or students' perceptions of the impacts of classroom assessment.

Phye (1997) states that "it is not only the assessment option that determines what we get as evidence of learning or achievement. How we use the assessment instruments or techniques also determine the nature of the knowledge a student is demonstrating. *How* we assess determines *what* we get" and thus classroom learning and assessment "go hand in hand" (p.51).

Airasian (1984) reviews literature that suggests teachers focus their classroom assessments in two areas: academic achievement and social behavior. The importance of these factors varies with grade level, with elementary teachers placing greater importance on social behavior. Airasian also found that teachers' informal "sizing up" assessments remain relatively stable throughout the year and influence student self-perceptions of ability.

Fleming and Chambers (1983), in a study that analyzed nearly 400 teacher-developed classroom tests, came to several conclusions:

- Short-answer questions are used most frequently.
- Essay questions are avoided, representing slightly more than 1% of test items.
- Matching items are used more than multiple choice or true false items.
- Most test questions, approximately 80%, sample knowledge of terms, facts, and rules and principles (94% for middle school teachers, 69% for high school teachers, and 69% of elementary school teachers).
- Few test items measure student ability to apply what they have learned.

Research by Carter (1984), in which the test development skills of high school teachers were studied, in support of what Fleming and Chambers found, reported that the teachers had considerable difficulty recognizing or writing items that tapped "higher order" thinking skills, such as application. Stiggins and Conklin (1992), with a sample of thirty-six teachers, found that recall knowledge items were used approximately fifty percent of the time.

There is ample evidence to suggest that many teachers do not have sufficient knowledge and skill to develop, apply, and summarize classroom assessments. In a survey of 228 teachers from four grades (2, 5, 8, and 11), Stiggins and Conklin (1992) report that nearly three fourths of the teachers indicated some concern about their own tests. Examples of the kinds of concerns expressed included: "Are my tests effective? How can I make them better? Do they focus on students' real skills? Are they challenging enough? Do they aid in learning?" (p. 39). Concern was greatest for high school teachers. Only 15% of high school teachers indicated that they had no concerns about their assessments. Stiggins and Conklin also asked twenty four teachers to keep a journal to reflect upon their assessment practices. The analysis focused on how teachers describe their assessments and what specific issues were raised related to their assessments. They found that teachers were most interested in assessing student mastery or achievement, and that performance assessment was used frequently. Few teachers emphasized higher order thinking skills. Finally, Stiggins and Conklin observed four sixth grade teachers and found that classroom assessments were integrated with instruction, using the results to inform instructional decision-making. The nature of the assessments used in each class was coupled closely with the roles each teacher set for her students, teacher expectations, and the type of teacher-student interactions desired. The results of these investigations led to the development of classroom

assessment profiles. The profile was tested with eight high school classrooms, resulting in the following key factors:

- Assessment purposes
- Assessment methods
- Criteria used in selecting assessment methods
- Quality of assessments
- Feedback to students
- Teacher as assessor (background, preparation)
- Teacher perception of the students
- The assessment-policy environment

These components can be used to characterize diverse assessment practices and environments.

Two recent studies document teacher beliefs and knowledge about classroom assessment. Frary, Cross, and Weber (1993) used a statewide random sample of 536 high school teachers of academic subjects to survey self-report practices and beliefs about classroom assessment. Frequency of use of various kinds of test questions revealed the following percentages:

Type of Question	Seldom or never	Frequently or always
Short answer	17%	56%
Essay	41%	38%
Multiple choice	21%	52%
True-false	47%	19%
Performance	30%	37%

These results suggest that teachers use a variety of assessment approaches. The teachers were asked to indicate degree of agreement to many statements concerning grading and assessment practices. Concerning assessment, it was noteworthy that 66% of the teachers agreed that essay tests provide a better assessment of student knowledge than do multiple choice tests; that 47% agreed that the nature of multiple choice items encourages superficial learning; and that better

measurement occurs when teachers award partial credit rather than scoring simply right or wrong.

A second survey of teachers, taken in 1992, was structured to obtain teacher competency concerning assessment practices by asking teachers to indicate which of several possible answers to assessment questions was best (Plake and Impara, 1997). A national random sample of 555 elementary, middle, and high school teachers was used. Overall mean performance on the survey was 66% correct. Teachers did better on items related to choosing and administering assessments and significantly worse on communicating results. According to the authors, the results "give empirical evidence of the anticipated woefully low levels of assessment competency for teachers" (p.67). The results also showed that teachers who had had a measurement course performed better than teachers who lacked this background.

In summary, the small amount of existing literature on classroom assessment practices indicates that teachers probably need further training to improve the quality of the assessments that are used. There continues to be reliance on selected-response tests, with conflicting evidence concerning the use of essays. Whatever the type of question, few are written to tap students' higher level thinking skills. Appropriately, teachers appear to use a variety of assessment methods. There is clearly a need for more research on classroom assessments. Classroom assessments consume significant amounts of time for both teachers and students, and have important consequences. Particularly absent in the literature are examination of relationships between classroom assessment practices and grading, how teachers use assessments to set standards, and how teachers make decisions about the assessments they use.

III. Grading Practices

Teachers' grading practices have received far more attention in the literature than have assessment practices. This may be due to the salient and summative nature of grades to students and parents. Grades have important consequences and communicate student progress to parents.

A study by Stiggins, Frisbie, and Griswold (1989) set the stage for research on grading by providing an analysis of current grading practices as related to recommendations of measurement specialists and newly established Standards for Teacher Competence in Educational Assessment of Students (American Federation of Teachers, National Council on Measurement in Education, National Education Association, 1990). In this study the authors interviewed and/or observed 15 teachers on 19 recommendations from the measurement literature. They found that teachers use a wide variety of approaches to grading, and that they wanted their grades to both fairly reflect student effort and achievement, as well as to motivate students. Contrary to recommended practice, it was found that teachers value student motivation and effort, and set different levels of expectation based on student ability. The authors recommended a research agenda in the following six areas to respond to these issues:

Area Needing Research	Illustrative Research Questions
Nature and Role of Nonachievement Factors	How do teachers define such traits as ability and effort? How do they assess these traits? Specifically, how are these traits factored into grades? What happens if these factors are reported separately?
Grades and Motivation	How does level of effort relate to the actual level of achievement? What role do grading practices play in causing students to set their own academic expectations of themselves? What role do grading practices play in causing students to give up and drop out?
Nature and Quality of Data Sources	How do homework completion records and homework performance data relate to scores on major tests over the same material? How reliable are scores achieved on teacher developed tests and how reliable are composite achievement indexes formed by aggregating those scores? How reliable are scores achieved on homework assignments?
Grade Computation Strategies	What effects do various misapplications of component weights have on the distribution of composite scores and grades? When teachers use percentage cutoff scores applied to achievement averages to determine grades, how do they account for variation in test difficulty? What do teachers understand a borderline average to mean? How do they resolve it?
Grading Policies	Are current grading policies consistent with sound practice? Do teachers know, understand, and implement policies?
Grade Interpretation	What do teachers understand grades to mean? How do they interpret the previous grades of students? What decisions do they make on the basis of a grade? How do students interpret grades? How do parents interpret grades?

Brookhart (1994) conducted a comprehensive review of literature on teachers' grading practices. Her review identified 19 studies completed since 1984. Seven studies investigated secondary school grading, 11 studies both elementary and secondary, and one study elementary teachers. Three general methods of study were identified: surveys in which teachers responded to questions concerning components included in grading, grade distributions, and attitudes toward grading issues; surveys in which teachers were asked to respond to grading scenarios, asking what they would do in various circumstances; and qualitative methods, including interviews, observation, and document analysis. Despite methodological and grade level differences, the findings from these studies are remarkably similar. This suggests that conclusions warranted from the research are generalizable. Taken together, Brookhart comes to the following conclusions:

- Teachers inform students of the components used in grading.
- Teachers try hard to be fair in grading.
- Measures of achievement, especially tests, are major contributors to grades.
- Student effort and ability are used widely as components of grades.
- Elementary teachers rely on more informal evidence and observation, while secondary teachers use paper and pencil achievement tests and other written evidence as major contributors.
- Teachers' grading practices vary considerably from one teacher to another, especially in perceived meaning and purpose of grades, and how nonachievement factors will be considered.
- Teachers' grading practices are not consistent with recommendations of measurement specialists, especially confounding effort with achievement.

In one study, Brookhart (1993) investigated the meaning teachers give to grades and extent to which value judgments are used in assigning grades. She used a sample of 84 teachers from all grade levels. Each teacher read seven scenarios about grading with multiple choices for responses about what the teacher would do in each situation. This was followed by an open-

ended question in which teachers explained the reasons for their choice. The results indicated that low ability students who tried hard would be given a passing grade even if the numerical grade were failure, while working below ability level did not affect the numerical grade. That is, an average or above average students would get the grade earned, whereas a below average student gets a break if there is sufficient effort to justify it. Teachers were divided about how to factor in missing work. About half indicated that a zero should be given, even if that meant a failure for the semester. The remaining teachers would lower the grade but not to a failure. The teachers' written comments showed that they strived to be "fair" to students. This sense of justice for all students was reflected in statements like "If grading criteria are clearly known by students, they should be followed," and "When questioned about a grade, I can show I was fair to all the students" (p.136). Teachers also seemed to indicate that a grade was a form of payment to students for work completed. More comments indicated that grades were something students earned as compared to grades indicating academic achievement, as compensation for work completed. This suggests that teachers, either formally or informally, include conceptions of student effort in assigning grades. Because teachers are concerned with student motivation, self-esteem, and the social consequences of giving grades, using student achievement as the sole criteria for determining grades is rare. This is consistent with earlier work by Brookhart (1991), in which she pointed out that grading often consists of a "hodgepodge" of attitude, effort, and achievement.

Cross and Frary (1996) report similar findings concerning the "hodgepodge" nature of grades. They surveyed 310 middle and high school teachers of academic subjects in a single system as well as 7367 students from the same system. A teacher survey was used to describe grading

practices and opinions regarding assessment and grading. The student survey asked about perceived importance teachers give to various factors and their satisfaction with the grading process. Consistent with Brookhart, it was reported that 72% of the teachers raised the grades of low ability students. A majority of students (55%) agreed that to be fair student ability should be considered. One-fourth of the teachers indicated that they raise grades for high effort "fairly often." One-third of the students indicated that their teachers considered effort. Almost 40% of the teachers indicated that student conduct and attitude were taken into consideration when assigning grades. A substantial majority of students (71%) endorsed the use of conduct and attitudes for determining grades. Interestingly, a very high percentage of teachers and students (81% and 70%, respectively) agreed that effort and conduct should be reported separately from achievement. Over half of the teachers reported that class participation was rated as having a moderate or strong influence on grades.

An earlier statewide study by Frary, Cross, and Weber (1993), using the same teacher survey that was used by Cross and Frary (1996), found similar results. Percentages of teachers agreeing or tending to agree to the following statements illustrates this conclusion:

Item	Percentage
• A student's ability should be taken into consideration in awarding final grades.	66
• An exceptionally low or high degree of student effort should be recognized by adjustment of the final grade.	66
• The amount of knowledge a student *gains* over the instructional period should be taken into consideration in awarding the final grade.	85
• Laudatory or disruptive classroom behavior should be considered in determining final grades.	31
• The minimum passing score on a test should be based at least in part on the scores earned by students of marginal ability who have be been putting forth satisfactory effort.	64

Another recent study by Truog and Friedman (1996), further confirms the notion of hodgepodge grading. In their study the written grading policies of 53 high school teachers were analyzed in relation to grading practices recommended by measurement specialists, and a focus group of eight teachers was conducted to probe reasoning used by the teachers. The study was based on an earlier investigation by Stiggins, Frisbie, and Griswold (1989) which found discrepancies between grading practices of teachers and recommended practice on 11 of 15 grading procedures and policies, including the use of effort and other nonachievement factors. Friedman and Manley (1991) also found that teachers routinely use ability, attitude, effort, participation, and other factors in addition to achievement when determining grades. Truog and Frieman (1996) found that written policies were consistent with earlier studies of teacher beliefs and practice. Nine percent of the teachers included ability as a factor in determining grades, 17% included attitude, 9% included effort, 43% included attendance, and 32% included student behavior.

Another survey of 143 elementary and secondary school teachers conducted by Cizek, Fitzgerald and Rachor (1995) collected data on teachers' assessment-related practices. Results indicated that assessment practices "were highly variable and unpredictable from characteristics such as

practice setting, gender, years of experience, grade level or familiarity with assessment policies in their school district" (p. 159). Furthermore, teachers generally use a variety of objective and subjective factors to maximize the likelihood that students obtain good grades. Overall, the authors concluded that "many teachers seemed to have individual assessment policies that reflected their own individualistic values and beliefs about teaching" (p.160). The authors argue that grades should be used in more meaningful ways to communicate about student performance.

In summary, the literature on grading strongly supports the notion that teachers believe it is important to combine nonachievement factors, such as effort, ability, and conduct, with student achievement to determine grades. While the studies are clear in this conclusion, less is known about how teachers decide to weigh these nonachievement factors in determining grades. Also, many of the surveys and other approaches in previous studies have asked teachers about their beliefs or projected behavior based on scenarios. It is possible that actual grading practice may be different. Despite increased focus on assessment and teacher competence with respect to measurement and grading, there appears to be a continuing discrepancy between recommended practice and teacher beliefs about grading. Furthermore, while descriptions of grading practices are plentiful, there is little research on the relationship between grading practices and student motivation and achievement. The fourth area of review represents an initial series of investigations of these relationships.

IV. Effect of Classroom Assessment and Grading Practices on Student Learning and Motivation

While there is little empirical literature on the effect of assessment and grading practices on student learning and motivation, Brookhart (1997) has recently suggested a theory about the role

of classroom assessment in motivating students. Her theory is based on a synthesis of classroom assessment literature and social cognitive theories of motivation. Social cognitive theories of motivation are based on the idea that perceptions and beliefs are central to the effect of environmental stimuli on motivation (Stipek, 1998). As students actively process assessment events they develop cognitions concerning task importance or value, difficulty, and the likelihood of success. These beliefs, in turn, influence expectations, effort, and motivation. Brookhart (1997) has depicted her theory graphically by showing how instruction, perceived task characteristics, and perceived self-efficacy influence effort, which in turn influences achievement (Figure 1).

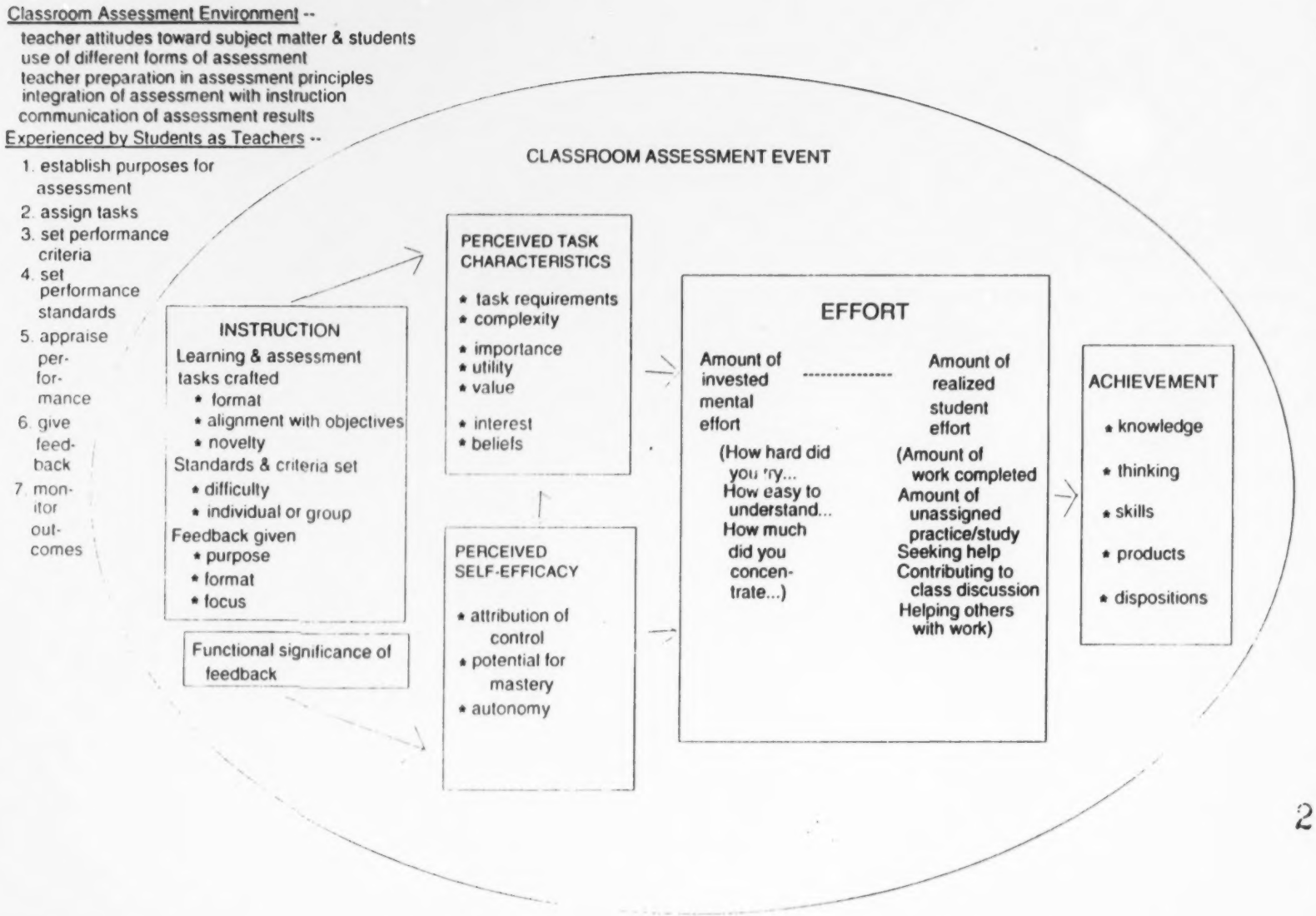

Figure 1. Model of a theoretical framework for investigating the effects of classroom assessment on student effort and achievement.

Pintrich and Schrauben (1992) point out that an important component of student effort is the perceived importance, utility, and value of engaging in the task. This is determined in part on why it is important to engage in the task. If students believe it is important for accomplishing future goals or because it has intrinsic interest, they will be more engaged. Intrinsic interest is established if the assessment task is challenging or raises curiosity, or is related to every day living. Several researchers have established goal orientation as an important component of motivation (Ames, 1992; Pintrich & Schunk, 1996; Pintrich & Schrauben, 1992). When students' goal orientation is mastery they are concerned most with developing new skills and acquiring new knowledge. Mastery orientations are more intrinsic. There is enjoyment, challenge, and meaningfulness in the task. A mastery orientation is related to more positive attitudes, use of effective learning strategies, and a belief that effort would lead to success. In contrast, a performance orientation results in students being motivated by achieving for success, such as a good grade or by performing better than other students. Rewards are usually extrinsic. Students are concerned most with what grade is achieved rather than what is learned. While these findings have been shown to hold for instructional tasks, it is reasonable to postulate, as Brookhart (1997) does, that assessment activities are framed and administered, as a task, to influence importance, utility, value, and goal orientation.

It is well established that self-efficacy is strongly related to student motivation (Pintrich & Schunk, 1996; Schunk, 1994). Self-efficacy is a student's self-conception of their ability to perform well on a specific task, to master the material, accomplish the task, or perform the skill. Self-efficacy helps to determine persistence and how hard students will try. From the standpoint of assessment, self-efficacy is affected by characteristics of the test or required performance. If

the assessment task is viewed as too difficult, students will not have a strong self-efficacy because they will tend to believe that they won't be able to do well. Self-efficacy is determined in part by knowing the nature of the scoring or the criteria upon which the performance will be judged. If scoring criteria or item difficulty are unknown, then there is little basis to support a strong sense of self-efficacy. However, when students know in advance how they will be judged by knowing the scoring criteria and by seeing examples of test items, papers, or other demonstrations of performance that have been graded, they are better able to connect the requirements to specific actions they can take to show achievement. Making this connection enhances self-efficacy because students are able to discern what, specifically, needs to be done.

Self-efficacy is also affected by student attributions. Attributions are the reasons students give themselves to explain why they performed as they did. They are the causal determinants of their performance (Pintrich & Schunk, 1996). Some attributions, such as ability and effort, are internal; others, such as task difficulty, teachers, and luck, are external. Attributions vary in the extent to which they are controllable. For example, effort and cheating are controllable, but ability is not; sometimes teachers can be controlled, but luck is clearly not controlled. Finally, stability of attributions can differ. For example, ability, retaining the same teacher, and overall ability of the class would be stable, whereas effort, luck and health are unstable. It has been demonstrated that if students' attributions following success are internal and stable or controllable, self-efficacy will be enhanced (Weiner, 1985). That is, if students believe that they did well because they tried, rather than because the test was easy, they will develop a strong self-efficacy with an expectation for continued success when required to perform similar tasks. On

the other hand, if success is ascribed to luck or some other external determinant, self-efficacy may remain low.

What is relevant about attributions for classroom assessment is how the assessment task and teacher feedback following performance affects the nature of the attributions that are formed. If the task is viewed as too difficult or too easy it will encourage external attributions. Tests and other assessments that are viewed as moderately difficult are less likely to be attributed externally. Feedback from teachers can take many forms, each with the potential of providing powerful messages to students about their level of effort and ability. Students appear to be especially vulnerable to teacher feedback about their ability. When students do poorly, any hint that the reason is due to low ability is likely to be endorsed, lowing self-efficacy. It is better to help the student attribute poor performance either to low effort, which is controllable (as long as the student did indeed give low effort), or to specific skills and knowledge that can be learned (unstable factors). For success, it is important to give feedback that establishes moderate effort and ability as attributions. Of course, this can't be credible unless, in fact, the student has engaged in a moderate level of effort. When grades or comments are vague and general, e.g., "well done" or "good job," the feedback is not likely to have much effect on self-efficacy. Students need help in drawing linkages between their performance and how and what they studied, and this is best accomplished with specific, individualized feedback.

Based on Brookhart's theory and other motivational literature, general effects of different assessment and grading practices can be expected, as illustrated in the following examples:

Assessment or Grading Practice	Effect on Motivation
Using grades as extrinsic rewards for desired and punishment for undesirable behavior.	Decreases motivation by focusing attention on performance goals rather than mastery goals; engenders feelings of being controlled; mitigates intrinsic motivation.
Being clear about how learning will be evaluated.	Enhances motivation by allowing students to self-check learning. Decreases anxiety of unknown evaluation.
Providing specific feedback following an assessment activity.	Enhances motivation by showing the link between effort and achievement, which strengthens self-expectations, and by helping students understand what needs to be changed to improve.
Extensive use of matching and fill-in-the-blank items.	Decreases motivation by emphasizing surface meaning and rote memorization.
Overly specific and focused test items.	Decreases motivation by narrowing preparation.
Using mistakes to show students how learning can be improved.	Enhances motivation by mitigating fear of failure.
Grading on the curve.	Decreases motivation for some students by emphasizing competition among students for scarce rewards, by focusing on performance goals (grades) rather than mastery, and by emphasizing external attributions.
Using very hard or very easy tests.	Decreases motivation by removing challenge.
Using moderately difficult tests.	Enhances motivation by providing some challenge and an exercise that will provide meaningful feedback; encourages internal, controllable attributions.
Use many assessments rather than a few major tests.	Enhances motivation by mitigating test anxiety and fear of failure.
Use of zeros in calculating grades for work not completed.	Decreases motivation if zeros make future performance meaningless.

Use of "authentic" assessment tasks.	Enhances motivation by connecting assessments to real life activities or situations, increasing importance, utility, and value.
Use of tests and quizzes to control student behavior.	Decreases motivation by limiting self-determination and intrinsic interest.
Use of pre-established criteria for evaluating student work.	Enhances motivation by emphasizing effort attributions.
Give students good grades for participation.	Decreases motivation by undermining intrinsic interest.
Provide incremental feedback.	Enhances motivation by increasing self-efficacy.
Provide public scoring criteria prior to administering the assessment task.	Enhances motivation by increasing self-efficacy. Public criteria help students know what to study and learn. Not using public criteria leads to a guessing game between students and teachers, resulting in little sense of self-efficacy.

Summary and Implications

The literature reviewed on the nature and effect of assessment and grading practices on student achievement has demonstrated that there is little empirical evidence of the specific effects of using particular assessments and grading procedures. This is due in part to the complex nature of teaching, and how assessment and grading are only a part of instruction. Assessment and grading continue to be a private activity, with considerable variation among teachers. While "newer" forms of assessment, such as performance-based and portfolio, are based on recent research on cognitive learning, the suggestions are based on theory and not empirical evidence. There are several studies which show that teachers engage in assessment and grading practices that are not consistent with what would be recommended by measurement "experts." For

example, combining nonachievement factors like effort, ability, and conduct with student achievement to determine grades, as well as "hodgepodge" grading. While descriptions of grading practices are plentiful, there is little research on the relationship between grading practices and student motivation and achievement. One theoretical model postulated by Brookhart (1997) represents an initial perspective about how assessment and grading practices affect self-efficacy, effort, and achievement. There is a strong research base with respect to the two major contributors to motivation (self-efficacy and importance, utility, and value), but not much about how specific assessment and grading practices effect these two components.

Brookhart's theory is reformulated in Figure 2 to provide more focus on the contributions of different assessment and grading procedures to each motivational component, and, subsequently, student performance. Working back from student performance, motivation, engagement, and effort is theoretically determined by three factors: student self-efficacy, student perception of assessment task importance, utility, and value, and type of assessment. The first two components are taken directly from the motivation literature. Type of assessment is added because it is well understood that this single factor, e.g., whether the test is multiple choice, essay, or performance-based, directly affects motivation. For example, we know that performance-based assessments are typically more engaging for students, and students study differently for essay tests than for objective tests. The type of assessment also influences perceived task value and task difficulty. For example, essay tests are usually viewed as more difficult than objective tests, and performance-based assessments usually have more value because they are typically based on problem solving in authentic contexts.

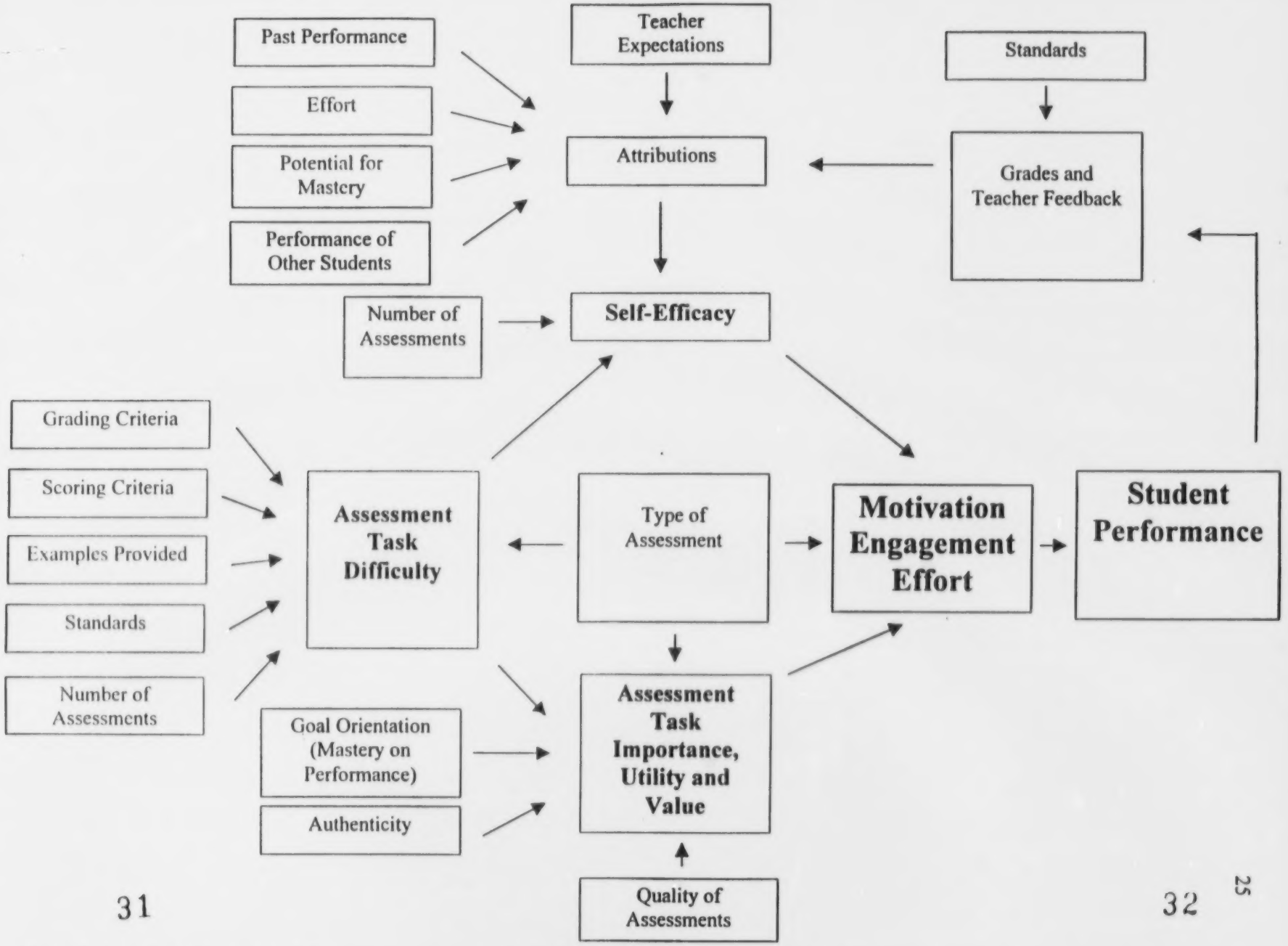

Figure 2. A model of the effect of classroom assessment and grading practices on student performance.

Assessment task importance, utility, and value are also influenced by goal orientation (mastery goals are more intrinsic and have more value than extrinsic performance goals), degree of authenticity, relevance or interest, and by the quality of the assessment. It is unlikely that students will perceive assessments of low quality as being important (e.g., unfair because of bias, not tied closed to instruction, little opportunity to learn, ambiguous learning targets). Perceived task difficulty is viewed as effecting both self-efficacy and assessment task importance, utility, and value. It is influenced by the nature of grading criteria, scoring criteria for individual assessments, whether examples of previously graded work are provided, the number of assessments given, and standards communicated to the students. When grading and scoring criteria are clear, fair and provided to students at the beginning of instruction, with examples of previous student work, when teacher standards are clear, and when there are many assessments, students will more likely perceive task difficulty as something that is within their ability.

Self-efficacy is viewed as being developed primarily from student attributions. These attributions are influenced by grades and teacher feedback, teacher expectations, past performance, effort expended, potential for mastery, performance of other students, and perceived task difficulty. Grades and teacher feedback are based on student performance and standards for performance. As previously noted, specific, individualized feedback lead to internal and controllable attributions, which in turn enhance self-efficacy. When teachers have high standards and expectations attributions tend to be more controllable (e.g., teachers who "won't accept" anything less than mastery). Attributions are also effected by past performance on similar tasks, effort expended for that task, and the performance of other students (doing well when most of class does poorly suggests internal, stable attribution of ability, e.g., "I must be

good at this if most of the class did poorly."). The potential for mastery also contributes to attributions and self-efficacy by providing hope. Students need to believe that it is possible for them to succeed on the basis of their own effort and ability.

Like Brookhart (1997), this model is a way of organizing the different assessment components into a framework that makes sense for understanding how classroom assessment contributes to motivation and student achievement. Clearly there is much to be researched to determine the utility of this model and other models.

References

Airasian, P. W. (1997). Classroom assessment (3rd Edition). NY: McGraw-Hill.

Airasian, P. W. (1984). Classroom assessment and educational improvement. Paper presented at the conference Classroom Assessment: A Key to Educational Excellence, Northwest Regional Educational Laboratory, Portland, Oregon.

American Federation of Teachers, National Council on Measurement in Education, and National Education Association (AFT, NCME, NEA). (1990) Standards for teacher competence in educational measurement. Washington, DC:. Author.

Ames, C. (1992). Classrooms: Goals, structures, and student motivation. Journal of Educational Psychology, 84, 261-271.

Brookhart, S. M. (1997). A theoretical framework for the role of classroom assessment in motivating student effort and achievement. Applied Measurement in Education, 10, 161-180.

Brookhart, S. M. (1994). Teachers' grading: Practice and theory. Applied Measurement in Education, 7, 279-301.

Brookhart, S. M. (1993). Teachers' grading practices: Meaning and values. Journal of Educational Measurement, 30, 123-142.

Brookhart, S. M. (1991). Grading practices and validity. Educational Measurement: Issues and Practice, 10, 35-36.

Carter, K. (1984). Do teachers understand the principles for writing tests? Journal of Teacher Education, 35, 57-60.

Cizek, G. J. (1997). Learning, achievement, and assessment: Constructs at a crossroads. In G. D. Phye, Editor, Handbook of classroom assessment, NY: Academic Press.

Cizek, G. J., Fitzgerald, Shawn M., & Rachor, Robert E. (1995). Teachers' Assessment Practices: Preparation, Isolation and the Kitchen Sink. Educational Assessment, 3(2), 159-179.

Fleming, M. & Chambers, B. (1983). Teacher-made tests: Windows on the classroom. In W. E. Hathaway, ed., Testing in the schools. New directions for testing and measurement. San Francisco: Jossey-Bass.

Frary, R.B., Cross, L.H. & Weber, L.J. (1993). Testing and grading practices and opinions of secondary teachers of academic subjects: Implications for instruction in measurement. Educational Measurement: Issues and Practice 12(3), 23-30.

Friedman, S. J. & Manley, M. (1991). Grading practices in the secondary school: Perceptions of the stakeholders. Paper presented at the Annual Meeting of the National Council on Measurement in Education, Chicago.

Haertel, E., Ferrara, S., Korpi, M., & Prescott, B. (1984). Testing in secondary schools: Student perspectives. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

McMillan, J. H. (1997). Classroom assessment: Principles and practice for effective instruction. Boston: Allyn & Bacon.

Phye, G. D. (1997). Classroom assessment: A multidimensional perspective. In G. D. Phye, Editor, Handbook of classroom assessment, NY: Academic Press.

Plake, B. S., & Impara, J. C. (1997). Teacher assessment literacy: What do teachers know about assessment? In G. D. Phye, ed., Handbook of classroom assessment, NY: Academic Press.

Pintrich, P. R., & Schrauben, B. (1992). Students' motivational beliefs and their cognitive engagement in classroom academic tasks. In D. H. Shunk & J. L. Meece (Eds.), Student perceptions in the classroom (pp. 149-183). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Pintrich, P. R., & Schunk, D. H. (1996). Motivation in education: Theory, research, and applications, Englewood Cliffs, NJ: Prentice-Hall.

Popham, W. J. (1995). Classroom assessment: What teachers need to know, Boston: Allyn & Bacon.

Schafer, W. D. (1993). Assessment in teacher education. Theory into Practice, 32, 118-126.

Schunk, D. H. (1994). Self-regulation of self-efficacy and attributions in academic settings. In D. H. Schunk & B. J. Zimmerman (Eds.), Self-regulation of learning and performance: Issues and educational applications (pp. 75-99). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Shulman, L. S. (1980). Test design: A view from practice. In E. L. Baker and E. S. Quellmalz, (Eds.), Educational testing and evaluation. Los Angeles: Sage.

Stiggins, R. J. (1997). Student-centered classroom assessment. Upper Saddle River, NJ: Prentice-Hall.

Stiggins, R. J. & Conklin, N. F. (1992). In teacher's hands: Investigating the practices of classroom assessment, Albany, NY: State University of New York.

Stiggins, R. J., Frisbie, D. A., & Griswold, P. A. (1989). Inside high school: Building a research agenda. Educational Measurement: Issues and Practice, 8, 5-14.

Stipek, D. (1998). Motivation to learn: From theory to practice. 3rd ed., Boston: Allyn & Bacon.

Truog, A. L. & Friedman, S. J. (1996). Evaluating high school teachers' written grading policies from a measurement perspective. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New York.

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. Psychological Review, 92(4), 548-73.

Annotated Bibliography

American Federation of Teachers, National Council on Measurement in Education and National Education Association (AFT, NCME, NEA). (1990). Standards for Teacher Competence in Educational Assessment of Students. Washington, DC:.Author.
These standards are a collaborative effort between AFT, NCME and NEA and are intended for use as:
1.) A guide for teacher educators as they design and approve programs for teacher preparation.
2.) A self-assessment guide for teachers in identifying their needs for professional development in student assessment.
3.) A guide for workshop instructors as they design professional development experiences for in-service teachers.
4.) An impetus for educational measurement specialists and teacher trainers to conceptualize student assessment and teacher training in student assessment more broadly than has been the case in the past.

Brookhart, S.M. (1995, April). Effects of the Classroom Assessment Environment on Achievement in Mathematics and Science. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.
As the title suggests, this paper addresses the effects of the classroom assessment environment on achievement in mathematics and science. The study results also support the beginnings of a theory to explain how classroom assessment influences student achievement. Sample size problems limit its generalizability to the national population, but the results are nonetheless informative. Results indicate that the classroom assessment environment should play an important role in student achievement, and suggest more focused field studies on the subject for better understanding.

Brookhart, S.M. (1997). A Theoretical Framework for the Role of classroom Assessment in Motivating Student Effort and Achievement. Applied Measurement in Education, 10, 161-180.
This study presents a theory about the role of classroom assessment in motivating student effort and achievement. The theory postulates that in any particular class, the classroom assessment environment is played out in repeated classroom assessment events, activity segments with associated expectations and assessments. Within a classroom assessment event, a teacher communicates to students through assignments, activities, and feedback on performance, and students respond according to their perceptions of these learning opportunities and their perceived efficacy to accomplish the tasks. The classroom assessment environment concept offers a way to integrate individual (students) and group (a class with one teacher) aspects of effort and achievement into one theoretical framework. The classroom assessment event concept offers a mechanism for how curriculum, instructional activities, and assessments impact student effort and achievement.

Brookhart, S.M. (1994). Teachers' Grading: Practice and Theory. Applied Measurement in Education, 7(4), 279-301.
This article is organized into two parts: 1.) A review of the literature on teachers' grading practices, and 2.) A discussion of the findings about teachers' grading practices in light of evaluation and motivation theory. Studies of grading practices document a gap between current practice and measurement theory. This article begins the process of bridging the gap between the two by developing a theoretical base that incorporates student motivation, classroom management and measurement functions into the grading process.

Brookhart, S.M. (1993). Teachers' Grading Practices: Meaning and Values. Journal of Educational Measurement 30,123-142.
Classroom teachers do not always follow recommended grading practices. Why not? It is possible to conceptualize this question as a validity issue and ask whether teachers' concerns over the many uses of grades outweigh concerns about the interpretation of grades. The purpose of this study was to investigate the meaning classroom teachers associate with grades, the value judgments they make when considering grades, and whether the meaning or values associated with grades differed by whether teachers had measurement instruction. A sample of 84 teachers, 40 with and 44 without measurement instruction, responded to classroom grading scenarios in two ways - with multiple-choice responses indicating what they would do and with written responses to the question, "Why did you make this choice?" A coding scheme based on Messick's (1989a, 1989b) progressive matrix of facets of validity was used for quantitative and qualitative analyses of written responses. The meaning of grades is closely related to the idea of student work; grades are pay students earn for activities they perform. The relationship of this notion to classroom management should be investigated. Teachers do make value judgments when assigning grades and are especially concerned about being fair. Teachers also are concerned about the consequences of grade use, especially for developing student self-esteem and good attitudes toward future school work. Measurement instruction made very little difference, although it did reduce the amount of self-referenced grading reported.

Canady, R. L. and Hotchkiss, P. R. (1989). It's a Good Score! Just a Bad Grade. Phi Delta Kappan, September 1989, 68-71.
In this article, the authors argue that schools and teachers must shift their focus from sorting and selecting students to better teaching of and learning by students. Consequently, assessment and grading practices must also change to reflect this new focus; Adversarial and inequitable grading policies must cease and new practices that increase students' likelihood of success must prevail. Seven problematic grading practices are addressed and suggested alternatives are provided.

Cross, L. H. and Frary, R. B. (1996, April). Hodgepodge Grading: Endorsed by Students and Teachers Alike. Paper presented at the annual meeting of the National Council on Measurement in Education, New York.
Previous research clearly documents that teachers often award what Brookhart (1991) has referred to as a "hodgepodge grade of attitude, effort and achievement" (p.36). This paper reports on a survey of grading practices involving 310 middle and high school students from the same system. The results largely validate the findings of earlier studies. Substantial majorities of the teachers reported "hodgepodge" grading practices. More important, the students largely

confirmed and supported the hodgepodge grading practices reported by their teachers. These results are contrasted with grading practices widely recommended in measurement texts followed by a discussion of how measurement specialists may be missing the mark in their efforts to communicate their views to teachers, school administrators, and the general public.

Cizek, G.J., Fitzgerald, S.M. and Rachor, R.E. (1995). Teachers' assessment practices: preparation, isolation and the kitchen sink. Educational Assessment, 3(2), 159-179.
A sample of 143 midwestern elementary and secondary school teachers from a variety of practice settings responded to a survey and provided comments regarding their assessment practices. The purpose of the survey was to collect background (demographic) information on the teachers and information on several assessment-related practices, including frequency with which teachers assign routine class assignments, types of marks used to report student performance, frequency and grading of major assignments and tests, source of classroom tests, kinds of marks used, methods used to combine marks, meaning of grades, teachers' knowledge and perceptions regarding district grading policies, and teachers' awareness of the grading policies of their peers. Interviews with teachers provided additional insights into their practices. Results indicated that teachers' assessment practices were highly variable and unpredictable from characteristics such as practice setting, gender, years of experience, grade level, or familiarity with assessment policies in their school district. Teachers generally claim to consider and incorporate a variety of objective and subjective factors when assigning grades on assignments, assessments, and report cards, synthesizing diverse kinds of information about achievement in ways that tend to maximize the likelihood that students will achieve high grades. Only about one half of the teachers surveyed indicated that they were aware of their districts' policies on grading, most were not aware of the assessment practices of their colleagues. Many teachers seemed to have individual assessment policies that reflected their own individualistic values and beliefs about teaching. Recommendations for making grades more meaningful ways of communicating about student performance are suggested.

Frary, R.B., Cross, L. H. and Weber, L. J. (1993). Testing and Grading Practices and Opinions of Secondary Teachers of Academic Subjects: Implications for Instruction in Measurement. Educational Measurement: Issues and Practice 12(3), 23-30.

The purpose of this study was to 1.) Document the extent to which problematic opinions and practices are present in a large, representative sample of secondary academic teachers. 2.) Document and characterize the need for remediation or training in measurement. Study questions included:

A.) To what extent do teachers interpret test scores as representing the percentage of knowledge that a student has learned?
B.) How pervasive is the practice of assigning letter grades directly on the basis of percent-correct scores?
C.) To what extent do teachers appreciate the need for relatively difficult tests if the ranking function is to be optimally served?
D.) To what extent do teachers endorse or believe in the efficacy of district-wide percentage grading scales?

E.) To what extent do teachers endorse the use of factors other than achievement in determining course grades?
F.) How do teachers determine the minimum passing score for a test?

Results showed that secondary teachers produce percent correct tests the scores from which merely rank students rather than indicate percent of some body of learned knowledge. In writing such tests, teachers hope and plan for score ranges between 60% (or 70%) and 100%, thus undermining the potential of their tests to reliably rank students. In service recommendations include exposure and training in measurement practice.

Higher Education Research Institute (1996). The American Freshman: National Norms for Fall 1996. Report from the Higher Education Research Institute, UCLA Graduate School of Education and Information Studies, Los Angeles.
This report examines issues and trends related to college freshmen. High school "grade inflation" and its relationship to increasingly competitive college admissions is discussed

McTighe, J. and Ferrara, S. (1994, November). Assessing Learning in the Classroom. A Report from Professional Standards and Practice. Report from the National Education Association, Professional Standards and Practice, Washington, DC.
A variety of methods are examined that teachers from preschool to graduate school levels can use in assessing their students; The common principles underlying classroom assessment are explored. The first principle is that the primary purpose of classroom assessment is to inform teaching and improve learning. A second principle is that multiple sources of information are necessary when assessing learning in the classroom. A third principle of classroom assessment concerns validity, reliability, and fairness. Once these principles are accepted, the selection of particular assessment methods should be based on desired learning outcomes, the purpose of the assessment, and audience for which it is intended. Assessment approaches that might be used include selected response items of the sort presented in multiple-choice, true-false, and matching tests and for performance-based approaches that include constructed responses, product assessment, performances, and process-focused assessment. In addition to making choices about classroom assessment methods, teachers should consider options for evaluating student work and for communicating assessment results. Various scales and reporting processes are discussed.. An appendix contains a glossary of classroom assessment terms.

Plake, B. and Impara, J. (1997). Teacher Assessment Literacy: What Do Teachers Know About Assessment? In G.D. Phye (Ed), Handbook of Classroom Assessment. Learning, Adjustment and Achievement (pp.68). San Diego, CA: Academic Press.
This article, and the one that follows, describes the results of a national research survey designed to measure teacher competency levels in educational assessment. This particular article discusses the validation process of the survey instrument used to do the study, and presents a lengthier discussion of the research findings. See Plake and Impara (1993) below for more

Plake, B. and Impara, J. (1993). Assessment Competencies of Teachers: A National Survey. Educational Measurement: Issues and Practice, 12(4), 10-12.
This article describes the results of a national research survey designed to measure teacher competency levels in educational assessment. Utilizing the *Standards for Teacher Competence in Educational Assessment of Students*, a two-part assessment device was developed to assess teachers' knowledge of identified competency areas. In general, teachers performed best in the competency area of administering, scoring and interpreting test results. Poorest performance was in the area of communicating test results. Teachers with training in measurement techniques scored statistically better than those without training. Those who expressed comfort interpreting standardized scores also scored statistically better than those who expressed discomfort interpreting standardized scores.

Selleri, P., Carugati, F., and Scappini, E. (1994). What Marks Should I Give? A Model of the Organization of Teachers' Judgments of Their Pupils. European Journal of Psychology of Education, 10(1), 25-40.
The present study is devoted to the empirical endeavor of showing the structural characteristics of this claimed general dimension, its longitudinal consistency, and its causal influence on the first level organization of judgments. A content analysis of school reports of 77 Italian pupils, filled out by their own five teachers over five years of compulsory school (from 6 to 10 years) show seven major topics, which are used by the teachers for their year-scheduled evaluations. A Lisrel-based Two-Level model of the organization of judgments is then presented and discussed. This model is shown to be well held by teachers at the end of the first school form and it allows to predict the organization of their evaluations during third and fifth form, as well as final judgments of each form. This model is discussed in a social psychological framework, which underlines the role played by normative aims of the school programs and the evaluative everyday practices as major professional duties for teachers.

Stiggins, R. J., Frisbie, D.A., and Griswold, P.A. (1989). Inside High School Grading Practices: Building a Research Agenda, Educational Measurement: Issues and Practice, Educational Measurement: Issues and Practice, 8(2), 5-14.
This article calls for further, in-depth, examination of the body of knowledge called "grading", as well as further authentication of the grading principles, methods and practices utilized by teachers. Employing a case study methodology, the researchers attempted to understand the values and procedures underpinning grading practices of 15 High School teachers. Actual teacher grading practices were compared to recommended practices and discrepancies were noted. Because this was not a random sample, no inferences can be drawn from the study. Nonetheless, steps have been undertaken to disentangle the complex array of myth, tradition, uncertainty and procedures that characterize grading practice.

Tittle, C. K. (1994). Toward an Educational Psychology of Assessment for Teaching and Learning: Theories, Contexts, and Validation Arguments. Educational Psychologist, 29, 149-162.
A framework for an educational psychology of assessment for teaching and learning is proposed, consisting of three dimensions: epistemology and theories, the interpreter and user, and assessment characteristics. The dimension of interpreter and user is equal in importance to

theory and assessments, responsive to cognitive constructivism and the construction of meanings and beliefs, as held by teachers and students in practice contexts. Illustrations of the lines of inquiry and evidence that follow from this framework are given, drawing on research with teachers and using a particular assessment. Validation arguments for assessments in a practice-based context will be stronger when they are proactive and include evidence on the constructions of teachers and students and the meanings and use an assessment has for them in their educational situation.

Truog, A. and Friedman, S. (1996). Evaluating high school teachers' Written Grading policies from a Measurement Perspective. Paper presented at the annual meeting of the National Council on measurement in Education, New York.
In the past, information about the grading practices used by high school teachers has come from questionnaires filled-out by teachers or observations/interviews. In this study, the written grading policies used by teachers (N=53) from a high school in the upper Midwest were analyzed to determine the extent to which they matched the grading practices generally recommended by measurement specialists. In addition, a follow-up focus group of teachers from the same school (N=8) met to discuss the practical implications of recommended practice to a large degree. The focus group discussion revealed that some teachers grade the way they do because they are responding to the expectations of parents, students, and their jobs as teachers. It is concluded that those with backgrounds in measurement and evaluation should become much more involved in helping to resolve the conflict that seems to exist between classroom reality and best practice in grading.

Wright, R.G. (1994). Success for All: The Median is the Key.
In this article, the author argues that grading students by the median is more appropriate than using other measures of central tendency. The median is the statistically correct measure of grades since grades consist of ordinal data or numbers on a scale whose intervals are uncertain or inconsistent. The more commonly used mean assumes, incorrectly, that grades are interval or ration data that carry information and implications beyond simple rank order. Use of the mean thus penalizes students for a few stumbles and thus does not accurately reward hard work. Grading by the median corrects this error.

Zhang, Z. and Burry-Stock, J. (1997, March). Assessment Practices Inventory: A Multivariate Analysis of Teachers' Perceived Assessment Competency. Paper presented at the annual meeting of the National Council on measurement in Education, Chicago.
The study was intended to (1.) determine the psychometric properties and the subscales of a 67-item Assessment practices Inventory (API) and (2.) examine the effects of measurement training and teaching experience on teachers' perceived assessment competency. Data were collected from 311 teachers on the API. The reliability of the API was supported by a Cronbach alpha of .97. Construct validity of the AOI was examined using Rasch model and factor analyses. Based on the factor analysis, seven composite scores were formed on which a 2x3 MANOVA was conducted to examine the effects of measurement training and teaching experience on teachers' perceived competency in seven assessment categories. Multivariate interaction effects between measurement training and years of teaching were significant (p less than .05). Subsequent examination revealed significant multivariate simple effects of measurement training at four or more years of teaching in tow factor-analyzed assessment categories (p less than .01). Follow up

comparisons between the means indicated that among the teachers who had taught four or more years, those with measurement training believed they were more skilled than those without measurement training in two main assessment categories (p less than .001; p less than .05). Implication for measurement training is discussed.

Zhang, Z. and Burry-Stock, J. (1995, November). A Multivariate Analysis of Teachers' Perceived Assessment Competency as a Function of Measurement Training and Years of Teaching. Paper presented at the annual meeting of the Mid-South Educational Research Association, Biloxi, MS.

This study investigated inservice teachers' assessment competency as a function of measurement training and years of teaching. Data were collected from 311 teachers on a 67-item Assessment Practices Inventory. Seven composite scores were formed based on the underlying dimensions from a principal factor analysis. A 2x3 MANOVA was conducted to examine the effects of measurement training and teaching experience on teachers' perceived competency in the seven assessment categories as reflected in the composite scores. Multivariate interaction effects between measurement training and years of teaching were significant. Subsequent examination revealed significant multivariate simple effects of measurement training at four or more years of teaching in two factor-analyzed assessment categories. Follow up comparisons between the means indicated that among the teachers who had taught four or more years, those with measurement training scored significantly higher than those without measurement training on standardized test results interpretation, classroom statistics, and using assessment results in decision making. This group also scored significantly higher on performance assessment and information observation. Appendixes contain tables of data and description of seven standards for teacher competence of educational assessment of students.